Lec 5: Analyzing Linear Systems

15-369/669/769: Numerical Computing

Instructor: Minchen Li

Table of Content

Table of Content

Positive Definiteness and the Cholesky Factorization

Properties of A^TA

- Recall: solving the least-squares problem $A\mathbf{x} \approx \mathbf{b}$ is equivalent to solving $A^TA\mathbf{x} = A^T\mathbf{b}$.
- Regardless of A, the matrix A^TA is:
  - **Symmetric**: $(A^TA)^T = A^T(A^T)^T = A^TA$;
  - **Positive Semi-Definite (PSD)**: $\forall \mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^TA^TA\mathbf{x} = \|A\mathbf{x}\|^2 \geq 0$.

When a matrix is symmetric and positive definite, it is called Symmetric Positive Definite (SPD).

> **Definition 4.1** (Positive (Semi-)Definite)**.** A matrix $B \in \mathbb{R}^{n\times n}$ is positive semidefinite if for all $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^\top B\mathbf{x} \geq 0$. B is positive definite if $\mathbf{x}^\top B\mathbf{x} > 0$ whenever $\mathbf{x} \neq \mathbf{0}$.

- A^TA is **positive definite** and **invertible** when A's column vectors are linearly independent.

Positive Definiteness and the Cholesky Factorization

Block Matrix Notation

- To solve SPD systems, we would like to build faster solvers utilizing the special structure.
- For convenience, we will use the block matrix notation:
  - Suppose $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times k}$, $C \in \mathbb{R}^{p\times n}$, and $D \in \mathbb{R}^{p\times k}$, we could construct a larger matrix:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \mathbb{R}^{(m+p)\times(n+k)}.$$

  - The mechanisms of matrix algebra generally extend to this case, e.g.,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} E & F \\ G & H \end{pmatrix} = \begin{pmatrix} AE+BG & AF+BH \\ CE+DG & CF+DH \end{pmatrix}.$$

Positive Definiteness and the Cholesky Factorization

Writing SPD Matrix in Block Matrix Form

We can deconstruct the symmetric positive-definite matrix $C \in \mathbb{R}^{n \times n}$ as a block matrix:

$$C = \begin{pmatrix} c_{11} & \mathbf{v}^\top \\ \mathbf{v} & \tilde{C} \end{pmatrix}$$

where $c_{11} \in \mathbb{R}$, $\mathbf{v} \in \mathbb{R}^{n-1}$, and $\tilde{C} \in \mathbb{R}^{(n-1)\times(n-1)}$. The SPD structure of C provides the following observation:

$$\begin{aligned} 0 &< \mathbf{e}_1^\top C \mathbf{e}_1 \text{ since } C \text{ is positive definite and } \mathbf{e}_1 \neq \mathbf{0} \\ &= \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} c_{11} & \mathbf{v}^\top \\ \mathbf{v} & \tilde{C} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = c_{11} \end{aligned}$$

By the strict inequality in the first line, we do not have to use pivoting to guarantee that $c_{11} \neq 0$ in the first step of Gaussian elimination.

Positive Definiteness and the Cholesky Factorization

Gaussian Elimination on SPD Matrix

Continuing with Gaussian elimination, we can apply a forward-substitution matrix E of the form

$$E = \begin{pmatrix} 1/\sqrt{c_{11}} & \mathbf{0}^\top \\ \mathbf{r} & I_{(n-1)\times(n-1)} \end{pmatrix}.$$

Here, the vector $\mathbf{r} \in \mathbb{R}^{n-1}$ contains forward-substitution scaling factors satisfying $r_{i-1}c_{11} = -c_{i1}$. Unlike our original construction of Gaussian elimination, we scale row 1 by $1/\sqrt{c_{11}}$ for reasons that will become apparent shortly.

By design, after forward-substitution, the form of the product EC is:

$$EC = \begin{pmatrix} \sqrt{c_{11}} & \mathbf{v}^\top/\sqrt{c_{11}} \\ \mathbf{0} & D \end{pmatrix},$$

for some $D \in \mathbb{R}^{(n-1)\times(n-1)}$.

Positive Definiteness and the Cholesky Factorization

Eliminating the 1st Row and Column

Now, we diverge from the derivation of Gaussian elimination. Rather than moving on to the second row, to maintain symmetry, we post-multiply by $E^\top$ to obtain $ECE^\top$:

$$\begin{aligned} ECE^\top &= (EC)E^\top \\ &= \begin{pmatrix} \sqrt{c_{11}} & \mathbf{v}^\top/\sqrt{c_{11}} \\ \mathbf{0} & D \end{pmatrix} \begin{pmatrix} 1/\sqrt{c_{11}} & \mathbf{r}^\top \\ \mathbf{0} & I_{(n-1)\times(n-1)} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & D \end{pmatrix}. \end{aligned}$$

- $\sqrt{c_{11}}\mathbf{r}^T + \dfrac{1}{\sqrt{c_{11}}}\mathbf{v}^T = \mathbf{0}^T$ because we constructed $r_{i-1}c_{11} = -c_{i1} = -v_{i-1}$, and so $\mathbf{r}c_{11} = -\mathbf{v}$.
- We have eliminated the 1st row and col of C, and D is SPD (to be proved in your assignment).

Positive Definiteness and the Cholesky Factorization

Cholesky Factorization

We can repeat this process to eliminate all the rows and columns of C symmetrically. This method is *specific* to symmetric positive-definite matrices, since

- symmetry allowed us to apply the same E to both sides, and
- positive definiteness guaranteed that $c_{11} > 0$, thus implying that ${}^{1}/_{\sqrt{c_{11}}}$ exists.

Similar to LU factorization, we have obtained a factorization $C = LL^\top$ for a lower-triangular matrix L. This factorization is constructed by applying elimination matrices symmetrically using the process above, until we reach

$$E_k \cdots E_2 E_1 C E_1^\top E_2^\top \cdots E_k^\top = I_{n\times n}.$$

$L \coloneqq E_1^{-1} E_2^{-1} \cdots E_k^{-1}$. $\boxed{\text{The product } C = LL^\top \text{ is known as the \textit{Cholesky factorization} of } C.}$

Positive Definiteness and the Cholesky Factorization

Cholesky Factorization Example, Initial Step

Example 4.6 (Cholesky factorization, initial step)**.** As a concrete example, consider the following symmetric, positive definite matrix

$$C = \begin{pmatrix} 4 & -2 & 4 \\ -2 & 5 & -4 \\ 4 & -4 & 14 \end{pmatrix}.$$

We can eliminate the first column of C using the elimination matrix E_1 defined as:

$$E_1 = \begin{pmatrix} 1/2 & 0 & 0 \\ 1/2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \longrightarrow E_1 C = \begin{pmatrix} 2 & -1 & 2 \\ 0 & 4 & -2 \\ 0 & -2 & 10 \end{pmatrix}.$$

We chose the upper left element of E_1 to be $1/2 = 1/\sqrt{4} = 1/\sqrt{c_{11}}$. Following the construction above, we can post-multiply by $E_1^\top$ to obtain:

$$E_1 C E_1^\top = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & -2 \\ 0 & -2 & 10 \end{pmatrix}.$$

Positive Definiteness and the Cholesky Factorization

Cholesky Factorization Example, Remaining Steps

Example 4.7 (Cholesky factorization, remaining steps)**.** Continuing Example 4.6, we can eliminate the second row and column as follows:

$$E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 1/2 & 1 \end{pmatrix} \longrightarrow E_2(E_1 C E_1^\top)E_2^\top = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 9 \end{pmatrix}.$$

Rescaling brings the symmetric product to the identity matrix $I_{3\times 3}$:

$$E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/3 \end{pmatrix} \longrightarrow E_3(E_2 E_1 C E_1^\top E_2^\top)E_3^\top = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Hence, we have shown $E_3 E_2 E_1 C E_1^\top E_2^\top E_3^\top = I_{3\times 3}$. As above, define:

$$L = E_1^{-1} E_2^{-1} E_3^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 2 & 0 \\ 2 & -1 & 3 \end{pmatrix}.$$

Positive Definiteness and the Cholesky Factorization

Cholesky Factorization, Practical Properties

- Takes half the memory to store the factor L compared to the LU factorization.
- The product LL^T is symmetric and PSD regardless of L;
  - If we factored $C = LU$ but made rounding and other mistakes, in degenerate cases the computed product $C' \approx LU$ may no longer satisfy these criteria exactly.
- Code for Cholesky factorization can be very succinct.

Positive Definiteness and the Cholesky Factorization

Cholesky Factorization, Implementation

- Suppose we choose an arbitrary k ∈ {1, . . . , n} and write L in block form isolating the k-th row and column:

$$L = \begin{pmatrix} L_{11} & \mathbf{0} & 0 \\ \boldsymbol{\ell}_k^\top & \ell_{kk} & \mathbf{0}^\top \\ L_{31} & \boldsymbol{\ell}_k' & L_{33} \end{pmatrix}.$$

$$C = LL^\top = \begin{pmatrix} L_{11} & \mathbf{0} & 0 \\ \boldsymbol{\ell}_k^\top & \ell_{kk} & \mathbf{0}^\top \\ L_{31} & \boldsymbol{\ell}_k' & L_{33} \end{pmatrix} \begin{pmatrix} L_{11}^\top & \boldsymbol{\ell}_k & L_{31}^\top \\ \mathbf{0}^\top & \ell_{kk} & (\boldsymbol{\ell}_k')^\top \\ 0 & \mathbf{0} & L_{33}^\top \end{pmatrix}$$
$$= \begin{pmatrix} \times & \times & \times \\ \boldsymbol{\ell}_k^\top L_{11}^\top & \boldsymbol{\ell}_k^\top \boldsymbol{\ell}_k + \ell_{kk}^2 & \times \\ \times & \times & \times \end{pmatrix}.$$

- $L_{11}\boldsymbol{\ell}_k = \mathbf{c}_k$, where L_{11} is **lower-triangular** and already computed when processing the k-th row
  - Solve $\boldsymbol{\ell}_k$ via forward-substitution
- $c_{kk} = \boldsymbol{\ell}_k^T \boldsymbol{\ell}_k + \ell_{kk}^2$
  - Then calculate ℓ_{kk} and choose the positive value

Positive Definiteness and the Cholesky Factorization

Cholesky Factorization, Pseudo-Code

function CHOLESKY-FACTORIZATION(C)
  ▷ Factors $C = LL^T$, assuming C is symmetric and positive definite

  $L \leftarrow C$ ▷ This algorithm destructively replaces C with L
  for $k \leftarrow 1, 2, \ldots, n$
    ▷ Back-substitute to place $\boldsymbol{\ell}_k^\top$ at the beginning of row k
    for $i \leftarrow 1, \ldots, k-1$ ▷ Current element i of $\boldsymbol{\ell}_k$
      $s \leftarrow 0$
      ▷ Iterate over L_{11}; $j < i$, so the iteration maintains $L_{kj} = (\boldsymbol{\ell}_k)_j$.
      for $j \leftarrow 1, \ldots, i-1 : s \leftarrow s + L_{ij}L_{kj}$
      $L_{ki} \leftarrow {}^{(L_{ki}-s)}/_{L_{ii}}$

    ▷ Apply the formula for ℓ_{kk}
    $v \leftarrow 0$ ▷ For computing $\|\boldsymbol{\ell}_k\|_2^2$
    for $j \leftarrow 1, \ldots, k-1 : v \leftarrow v + L_{kj}^2$
    $L_{kk} \leftarrow \sqrt{L_{kk} - v}$
  return L

- Runs in $O(n^3)$ time;
- Takes around $\frac{n^3}{3}$ operations, half the work needed for LU.

Table of Content

Sparsity

Definition, Example, and Storage

- **Sparse matrix:** most of the entries are exactly zero, e.g.
  - **Image processing** systems link each pixel's value to its up/down/left/right neighbors. The system matrix $A \in \mathbb{R}^{p \times p}$ for p pixels is sparse with only $O(p)$ nonzeros per row.
  - **Machine learning:** Graphical models use nodes for variables and edges for dependencies. Linear systems have one row per node, with nonzeros only for that node and its neighbors.
- There is no reason to store n^2 entries of an $n \times n$ sparse matrix.
  - Sparse matrix storage techniques only store the $O(n)$ nonzeros in a more reasonable data structure, e.g., a list of row/column/value triplets.

Sparsity

Linear Solvers

- The LU (and Cholesky) factorizations of a sparse matrix A may not result in sparse L and U matrices;

- There are many direct sparse solvers that produce an LU-like factorization without inducing much additional nonzeros;

  - *T. Davis. Direct Methods for Sparse Linear Systems. Fundamentals of Algorithms. Society for Industrial and Applied Mathematics, 2006.*

- Alternatively, iterative techniques can obtain approximate solutions to linear systems using only multiplication by A and A^T.

Sparsity

Tridiagonal Matrix

Certain matrices are not only sparse but also *structured*. For instance, a *tridiagonal* system of linear equations has the following pattern of nonzero values:

$$\begin{pmatrix} \times & \times & & & \\ \times & \times & \times & & \\ & \times & \times & \times & \\ & & \times & \times & \times \\ & & & \times & \times \end{pmatrix}.$$

- **Remark:** Gaussian elimination provides only one option for solving linear system. It may be possible to show that the system matrix can be solved more easily by identifying special properties like symmetry, positive-definiteness, and sparsity.

  - *G. Golub and C. Van Loan. Matrix Computations. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2012.*

Table of Content

Sensitivity Analysis
Vector Norm

- Before we can discuss the sensitivity of a linear system, we need to define what it means for a change δx to be "small."

> **Definition 4.2** (Vector norm)**.** A vector norm is a function $\|\cdot\| : \mathbb{R}^n \to [0, \infty)$ satisfying the following conditions:
>
> - $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$ ("$\|\cdot\|$ separates points").
> - $\|c\mathbf{x}\| = |c|\|\mathbf{x}\|$ for all scalars $c \in \mathbb{R}$ and vectors $\mathbf{x} \in \mathbb{R}^n$ ("absolute scalability").
> - $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ("triangle inequality").

- e.g., 2-norm: $\|\mathbf{x}\|_2 \coloneqq \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$

Sensitivity Analysis
P-Norm

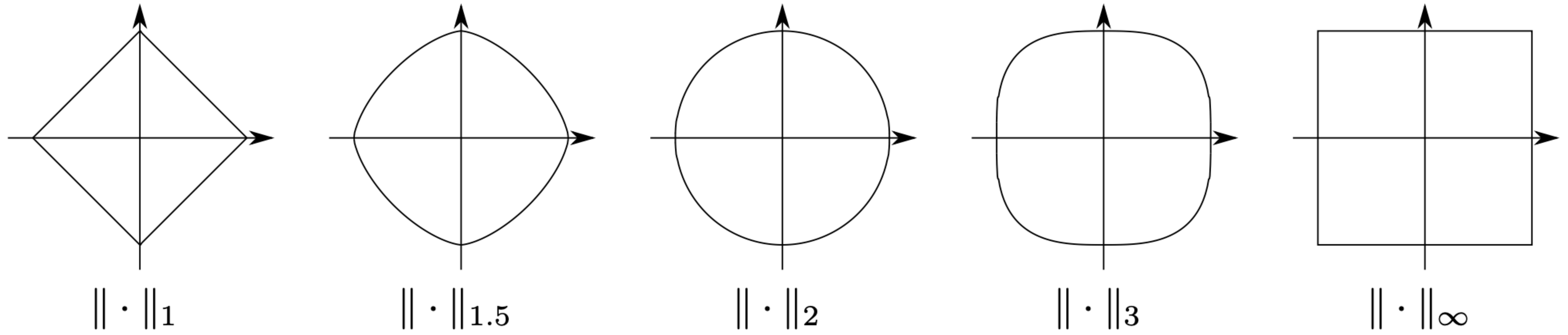


Figure 4.7 The set $\{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| = 1\}$ for different vector norms $\|\cdot\|$.

- $\|\mathbf{v}\|_p \leq \|\mathbf{v}\|_q$ when $p > q$

The p-norm $\|\mathbf{x}\|_p$, for $p \geq 1$, is given by

$$\|\mathbf{x}\|_p := \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p}.$$

Of particular importance is the 1-norm, also known as the "Manhattan" or "taxicab" norm:

$$\|\mathbf{x}\|_1 := \sum_{k=1}^{n} |x_k|.$$

The ∞-norm $\|\mathbf{x}\|_\infty$ is given by

$$\|\mathbf{x}\|_\infty := \max(|x_1|, |x_2|, \cdots, |x_n|).$$

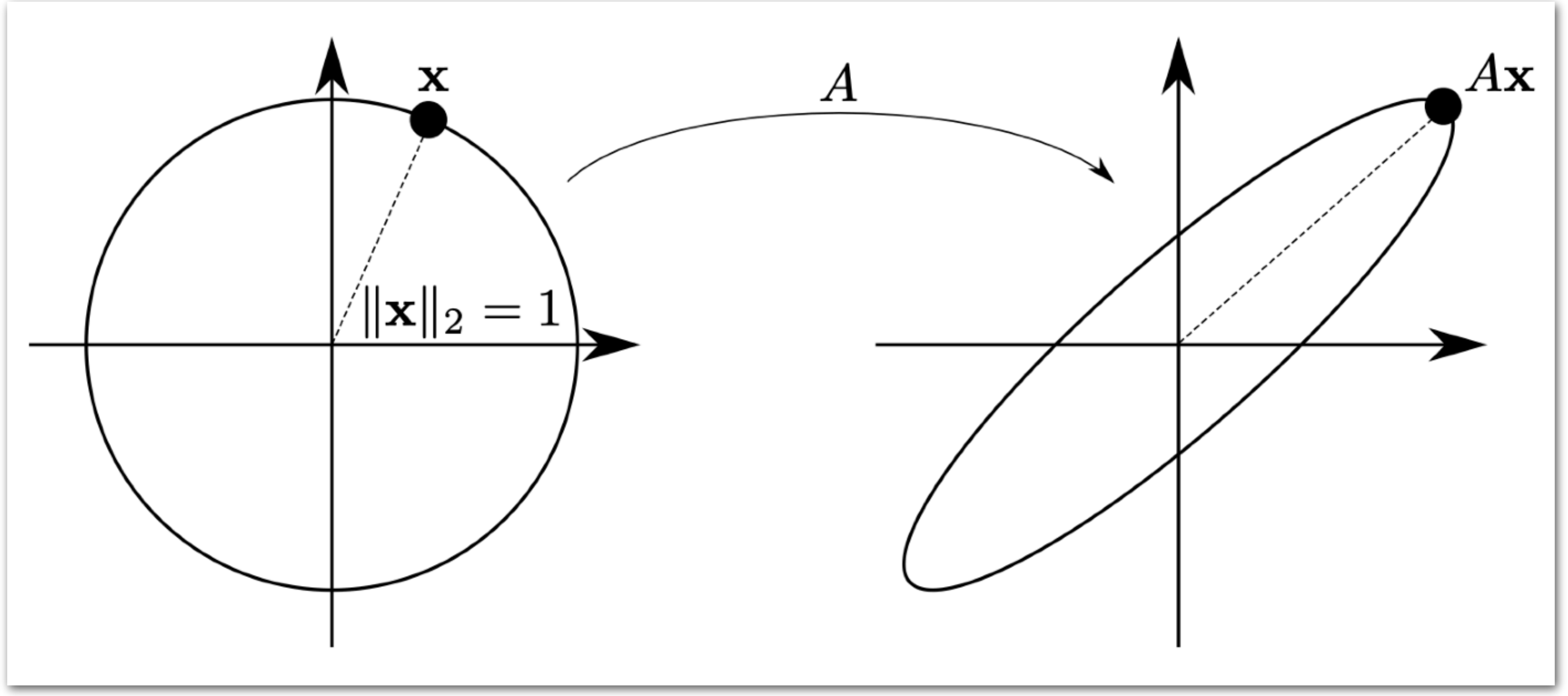
Sensitivity Analysis

Matrix Norm

- We can "unroll" any matrix in $\mathbb{R}^{m\times n}$ to a vector in $\mathbb{R}^{mn}$ to adapt any vector norm to matrices
  - e.g., Frobenius norm: $\|A\|_{\text{Fro}} := \sqrt{\sum_{i,j} a_{ij}^2}.$
  - But matrix norm constructed this way may not have a clear connection to the effect of $A\mathbf{x}$ for different vectors $\mathbf{x}$

> **Definition 4.4** (Induced norm)**.** The matrix norm on $\mathbb{R}^{m\times n}$ *induced* by a vector norm $\|\cdot\|$ is given by
>
> $$\|A\| := \max\{\|A\mathbf{x}\| : \|\mathbf{x}\| = 1\}.$$
>
> That is, the induced norm is the maximum length of the image of a unit vector multiplied by A.

Sensitivity Analysis
Induced 2-Norm



This definition in the case $\|\cdot\| = \|\cdot\|_2$ is illustrated in Figure 4.8. Since vector norms satisfy $\|c\mathbf{x}\| = |c|\|\mathbf{x}\|$, this definition is equivalent to the expression

$$\|A\| := \max_{\mathbf{x}\in\mathbb{R}^n\setminus\{0\}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}.$$

From this standpoint, the norm of A induced by $\|\cdot\|$ is the largest achievable ratio of the norm of $A\mathbf{x}$ relative to that of the input $\mathbf{x}$.

The induced two-norm, or *spectral norm*, of $A \in \mathbb{R}^{n\times n}$ is the square root of the largest eigenvalue of $A^\top A$. That is,

$$\|A\|_2^2 = \max\{\lambda : \text{there exists } \mathbf{x} \in \mathbb{R}^n\setminus\{\mathbf{0}\} \text{ with } A^\top A\mathbf{x} = \lambda\mathbf{x}\}.$$

Sensitivity Analysis

Condition Numbers

- Suppose we are given perturbation δA and $\delta \mathbf{b}$ to the linear system $A\mathbf{x} = \mathbf{b}$. For small ϵ, we can write a vector-valued function $\mathbf{x}(\epsilon)$as the solution to

$$(A + \varepsilon \cdot \delta A)\mathbf{x}(\varepsilon) = \mathbf{b} + \varepsilon \cdot \delta \mathbf{b}.$$

- Thus, we can expand the relative error made by solving the perturbed system:

$$\frac{\|\mathbf{x}(\varepsilon) - \mathbf{x}(0)\|}{\|\mathbf{x}(0)\|} \leq |\varepsilon| \underbrace{\|A^{-1}\|\|A\|}_{\kappa} \underbrace{\left(\frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|} + \frac{\|\delta A\|}{\|A\|}\right)}_{D} + O(\varepsilon^2)$$

> **Definition 4.5** (Matrix condition number)**.** The condition number of $A \in \mathbb{R}^{n \times n}$ with respect to a given matrix norm $\|\cdot\|$ is
>
> $$\operatorname{cond} A := \|A\|\|A^{-1}\|.$$
>
> If A is not invertible, we take $\operatorname{cond} A := \infty$.

Sensitivity Analysis

*Derivation of Condition Numbers

Differentiating both sides with respect to ε and applying the product rule shows:

$$\delta A \cdot \mathbf{x}(\varepsilon) + (A + \varepsilon \cdot \delta A)\frac{d\mathbf{x}(\varepsilon)}{d\varepsilon} = \delta \mathbf{b}.$$

Using the Taylor expansion, we can write

$$\mathbf{x}(\varepsilon) = \mathbf{x}(0) + \varepsilon \mathbf{x}'(0) + O(\varepsilon^2),$$

$$\begin{aligned}
\frac{\|\mathbf{x}(\varepsilon) - \mathbf{x}(0)\|}{\|\mathbf{x}(0)\|} &= \frac{\|\varepsilon \mathbf{x}'(0) + O(\varepsilon^2)\|}{\|\mathbf{x}(0)\|} \text{ by the Taylor expansion above} \\
&= \frac{\|\varepsilon A^{-1}(\delta \mathbf{b} - \delta A \cdot \mathbf{x}(0)) + O(\varepsilon^2)\|}{\|\mathbf{x}(0)\|} \text{ by the derivative we computed} \\
&\leq \frac{|\varepsilon|}{\|\mathbf{x}(0)\|}(\|A^{-1}\delta \mathbf{b}\| + \|A^{-1}\delta A \cdot \mathbf{x}(0))\|) + O(\varepsilon^2) \\
&\qquad \text{by the triangle inequality } \|A + B\| \leq \|A\| + \|B\| \\
&\leq |\varepsilon|\|A^{-1}\|\left(\frac{\|\delta \mathbf{b}\|}{\|\mathbf{x}(0)\|} + \|\delta A\|\right) + O(\varepsilon^2) \text{ by the identity } \|AB\| \leq \|A\|\|B\| \\
&= |\varepsilon|\|A^{-1}\|\|A\|\left(\frac{\|\delta \mathbf{b}\|}{\|A\|\|\mathbf{x}(0)\|} + \frac{\|\delta A\|}{\|A\|}\right) + O(\varepsilon^2) \\
&\leq |\varepsilon|\|A^{-1}\|\|A\|\left(\frac{\|\delta \mathbf{b}\|}{\|A\mathbf{x}(0)\|} + \frac{\|\delta A\|}{\|A\|}\right) + O(\varepsilon^2) \text{ since } \|A\mathbf{x}(0)\| \leq \|A\|\|\mathbf{x}(0)\| \\
&= |\varepsilon| \underbrace{\|A^{-1}\|\|A\|}_{\kappa} \underbrace{\left(\frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|} + \frac{\|\delta A\|}{\|A\|}\right)}_{D} + O(\varepsilon^2) \text{ since by definition } A\mathbf{x}(0) = \mathbf{b}.
\end{aligned}$$

Sensitivity Analysis
Properties of Condition Numbers

- For nearly any matrix norm, $\text{cond}\, A \geq 1$ for all A.

- Scaling A has no effect on its condition number.

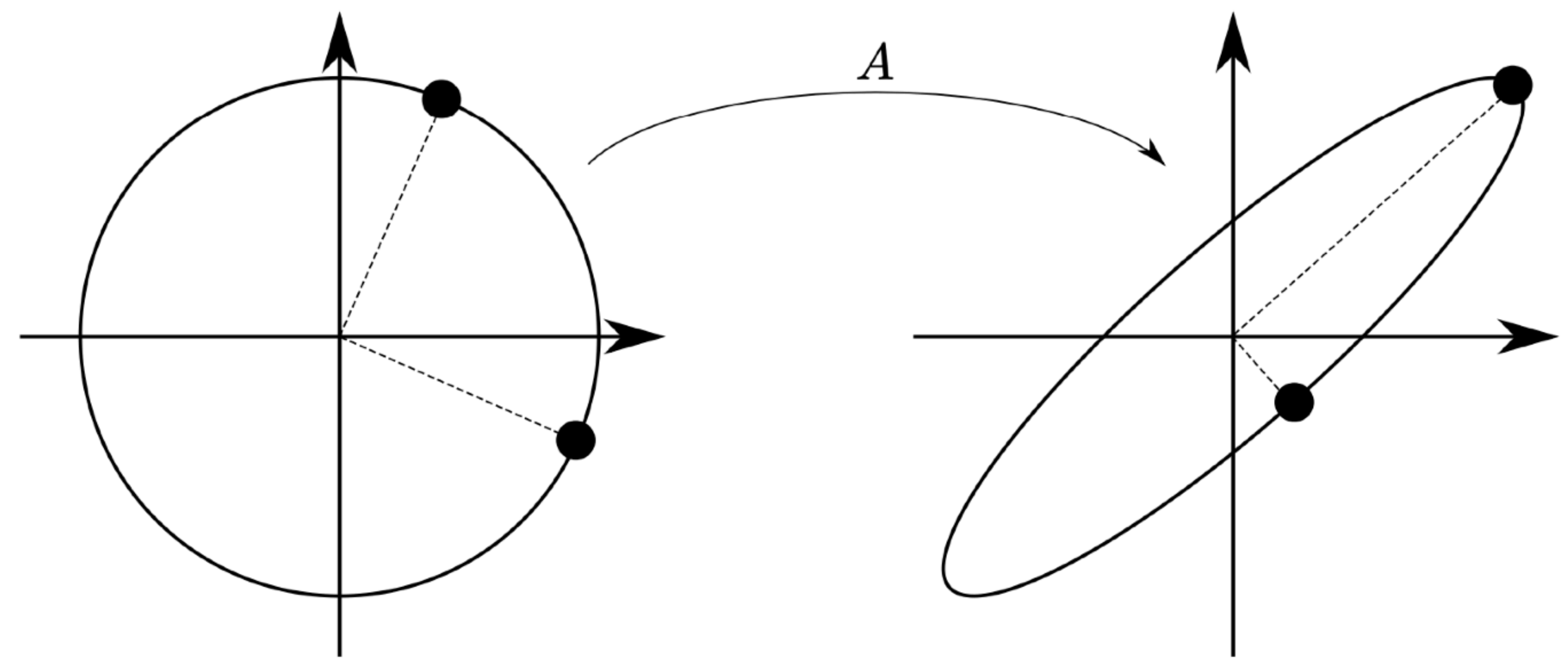


Figure 4.9 The condition number of A measures the ratio of the largest to smallest distortion of any two points on the unit circle mapped under A.

- Large $\text{cond}\, A$ indicate that solutions to $A\mathbf{x} = \mathbf{b}$ can be unstable under perturbations of A or b.

If $\|\cdot\|$ is induced by a vector norm and A is invertible,

$$\begin{aligned}
\|A^{-1}\| &= \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A^{-1}\mathbf{x}\|}{\|\mathbf{x}\|} \text{ by definition} \\
&= \max_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{y}\|}{\|A\mathbf{y}\|} \text{ by substituting } \mathbf{y} = A^{-1}\mathbf{x} \\
&= \left(\min_{\mathbf{y}\neq\mathbf{0}} \frac{\|A\mathbf{y}\|}{\|\mathbf{y}\|}\right)^{-1} \text{ by taking the reciprocal.}
\end{aligned}$$

$$\boxed{\text{cond}\ A = \left(\max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}\right)\left(\min_{\mathbf{y}\neq\mathbf{0}} \frac{\|A\mathbf{y}\|}{\|\mathbf{y}\|}\right)^{-1}.}$$

Table of Content